Spatial Assessment of Two Widely Used Land-Cover Datasets Over the Continental U.S.

Pei-Yu Chen, Mauro Di Luzio, and Jeff G. Arnold

Abstract*—Satellite-based land-cover datasets are required for various environmental studies. Two of the most widely used land-cover datasets for the U.S. are the National Land-Cover Data (NLCD) at 30-m resolution and the Global Land-Cover Characteristics (GLCC) at 1-km nominal resolution. Both datasets were produced around 1992 and were expected to contribute similar land-cover information. This study investigated the NLCD distribution within each of 11 GLCC classes at 1-km unit in ten U.S. states. Our analyses showed that the NLCD had similar spatial distribution as the GLCC for the classes of grassland, shrubland, as well as deciduous and evergreen forests. Meanwhile, the GLCC class of cropland and pasture was highly correlated to the NLCD classes of row crops and pasture/hay. The GLCC savanna was appropriately related to the NLCD grassland, pasture/hay, and deciduous forest. The NLCD classes of row crops, pasture, and deciduous forest mainly dominated the GLCC class of cropland/woodland mosaic. Spatial similarity was lower for the GLCC classes of mixed forest, wooded wetland, and cropland/grassland mosaic. In addition to the NLCD urban areas, the GLCC urban and built-up lands were consistently related to the NLCD vegetated areas due to the common mixture of urban and vegetated lands. A set of subclass land-cover information provided through this study is valuable to understand the degrees of spatial similarity for the global vegetated and urban-related classes in selected study areas. The results of this study provide great reference for interchanging less-detailed global land-cover datasets for detailed NLCD to support environmental studies.

***Index Terms*—Global Land-Cover Characteristics (GLCC), National Land-Cover Data (NLCD).**

I. Introduction

LAND-COVER information has become essential and critical for environmental studies and land use planning. The appropriate use of this information is usually a consequence of the derived spatial scale. For local studies, fine-resolution data (e.g., one to few meters) are generally more appropriate than coarse-resolution data (e.g., one to few kilometers), which conversely play an important role in several global environmental and climatic studies.

In the U.S., an intermediate-scale dataset, National Land-Cover Data (NLCD) [1], is often used for regional- as well as national-scale investigations. The NLCD, developed based on 1992 satellite imagery at 30-m resolution, was consistently classified and seamlessly covers the entire country. It provides a suitable land-cover dataset to support national environmental assessments using watershed-based water quality models, such as the Soil and Water Assessment Tool (SWAT) [2] and Spatially Referenced Regressions on Watershed (SPARROW) [3]. A general concern with NLCD is the update of the information content that takes extensive time and tremendous processing effort. In fact, the new version of the datasets [4], based on the 2000 vintage Landsat data, is still under completion by several federal agencies forming the MultiResolution Land Characteristics Consortium (MRLC).

The need for satellite-based land-cover data becomes critical for near real-time analysis on a national scale. Over the last few years, several research institutes and government organizations have updated extended land-cover databases. For example, a new global land-cover database for the year 2000/2001 was developed by using Moderate Resolution Imaging Spectrometer (MODIS) data [5]. The Canada Center for Remote Sensing implemented SPOT/Vegetation data to produce a national land-cover database for Canada for the year 1998 [6]. All of these newly produced land-cover database were generated at 1-km resolution.

Previously, another land-cover dataset, the Global Land-Cover Characteristics (GLCC), was produced at 1-km nominal spatial resolution for the entire globe [7], [8]. The GLCC was developed using satellite data collected across 1992 through 1993, like for the NLCD. Although developed based on different satellite sensors with different goals and resolutions, the two datasets were expected to contribute similarly to the land-cover information over the conterminous U.S.

Different vegetated land-cover types have varied influences on water and energy cycles. The expansion of urban distributions alters climate and rainfall patterns. Both vegetated and urban-related areas were important for land-cover studies. The objective of this study was to analyze the spatial distribution (zonal and pixel based) of the NLCD and GLCC over the U.S. This study investigates the NLCD distribution within the selected 1-km $\times$ 1-km GLCC pixels to analyze the spatial relationship of land-cover information between the two datasets. The results will contribute to the knowledge of land-cover correlation between fine-resolution and coarse-resolution dataset. Investigating the degree and areas of agreement between the two datasets can provide background information for interchanging less-detailed land-cover maps in a large area or whenever appropriate.

Manuscript received March 23, 2005; revised June 6, 2005. This work was supported by the U.S. Department of Agriculture's Agricultural Research Service through the Specific Cooperate Agreement 58-6206-1-005.

P.-Y. Chen and M. Di Luzio are with the Blackland Research and Extension Center, Texas Agricultural Experiment Station, Temple, TX 76502 USA (e-mail: pchen@brc.tamus.edu).

J. G. Arnold is with the Grassland, Soil, and Water Research Laboratory, U.S. Department of Agriculture—Agricultural Research Service, Temple, TX 76502 USA.

Digital Object Identifier 10.1109/TGRS.2005.854308

TABLE I
CLASSIFICATION SCHEMES OF GLCC AND NLCD, RESPECTIVELY, BASED ON THE MODIFIED LAND-COVER CLASSES OF ANDERSON LEVEL I. THE NUMBERS IN PARENTHESES REPRESENT LAND-COVER CODES

Modified land-cover classes of Anderson Level I	GLCC classification scheme	NLCD classification scheme
Urban or Built-up Land (1)	Urban and Built-up Land (100)	Low Intensity Residential (21) High Intensity Residential (22) Commercial/Industrial/Transportation (23)
Agricultural Land (2)	Dryland Cropland and Pasture (211)	Orchards/Vineyard (61)
	Irrigated Cropland and Pasture (212)	Pasture/Hay (81)
	*Mixed Dryland/Irrigated Cropland and Pasture (213)	Row crops (82)
	Cropland/Grassland Mosaic (280)	Small Grains (83)
	Cropland/Woodland Mosaic (290)	Fallow (84)
		Urban/Recreational Grasses (85)
Rangeland (3)	Grassland (311)	Grasslands/Herbaceous (71)
	Shrubland (321)	Shrubland (51)
	*Mixed Shrubland/Grassland (330)	
	Savanna (332)	
Forest Land (4)	Deciduous Broadleaf Forest (411)	Deciduous Forest (41)
	*Deciduous Needleleaf Forest (412)	
	Evergreen Broadleaf Forest (421)	Evergreen Forest (42)
	Evergreen Needleleaf Forest (422)	
	Mixed Forest (430)	Mixed Forest (43)
Water (5)	#Water Bodies (500)	Open Water (11)
Wetland (6)	Wooded Wetland (610)	Woody Wetlands (91)
	*Herbaceous Wetland (620)	Emergent/Herbaceous Wetlands (92)
Barren Land (7)	#Barren or Sparsely vegetated (770)	Bare Rock (31) Quarries/Mines (32) Transitional (33)
Tundra (8)	*Wooded Tundra (810) *Herbaceous Tundra (820) *Bare Ground Tundra (830) *Mixed Tundra (850)	
Perennial Snow or Ice (9)	#Snow or Ice (900)	Perennial Ice/Snow (12)

Excluded in this study
* None or few pixels in U.S.

II. MATERIALS AND METHODS

The NLCD and GLCC datasets were generated based on an entire year of satellite images. Both leaves-off and leaves-on datasets were analyzed to distinguish vegetation types. The NLCD is a 21-class land-cover classification scheme resembling the well-established Anderson land-use/land-cover classification system [9] (Table I). The GLCC datasets were provided in seven different land-cover classification schemes, and the U.S. Geological Survey (USGS) land-use and land-cover classification in 24 classes was adopted for this study due to similarity to the Anderson classification scheme (Table I). Both land-cover datasets are available for free download at the USGS web site.

The positional accuracy of each pixel is important, since this study investigated the NLCD distribution within 1 km $\times$ 1 km GLCC pixels. Each Landsat Thematic Mapper (TM) image used to create the NLCD was terrain-corrected using digital terrain elevation data and georegistered using ground control points, resulting in a root mean square registration error of less than one pixel (30 m) [1]. The registration accuracy of GLCC data has not been officially published yet. The USGS web documentation [10] stated that the goal of positional accuracy is 1-km or less for the Advanced Very High Resolution Radiometer (AVHRR) images used to create the GLCC. The ground control points and hydrographic features were used for geometric registration. Digital elevation data were incorporated to reduce registration errors [10].

A. Land-Cover Datasets

The NLCD was produced based primarily on 1992 vintage Landsat-5 TM data and was provided on a state-by-state basis [1]. The TM multiband data were processed using an unsupervised clustering algorithm [11]. The resulting spectral clusters were resolved into one of 21 thematic classes using aerial photography and ground observations. Clusters that represented more than one land-cover category were identified using various ancillary datasets and models developed to split the confused clusters into the correct land-cover categories.

The GLCC datasets were derived from National Oceanic and Atmospheric Administration (NOAA) AVHRR images [12]. The concept of seasonal land-cover regions was applied to the GLCC creation to present the temporal and spatial patterns of vegetation. The AVHRR normalized difference vegetation index (NDVI) composites and other ancillary geographic data, including digital elevation model, ecoregion data, and country or regional-level vegetation and land-cover maps, were used to classify global land-cover data. An unsupervised clustering method was applied to multitemporal NDVI composites, and each land-cover cluster was classified based on ancillary data and expert interpretation. Postclassification was implemented using multisource geographic data to refine classes containing two or more disparate land-cover types [13].

Accuracy assessment of land-cover maps derived from satellite data has been an important concern of the remote sensing community [14]. The 1992 NLCD has accuracy ranging from 37% (central U.S.) to 69% (western coast) (http://edcwww.usgs.gov/programs/lccp/accuracy). Much of the classification error occurred among the NLCD classes that aggregate into a single Anderson level I class. For example, pasture/hay was often confused with row crops, and mixed forest with deciduous forest and evergreen forest. Other sources of disagreement were between the forest and agricultural classes, and between the forested wetland class and the forest classes. Moreover, mines were often confused with transitional, hay/pasture, and deciduous forest [1], [8].

For the GLCC datasets, the averaged classification accuracy was 59.4% [15]. The highest individual class accuracies occurred in the classes of evergreen broadleaf forests (78%) and barren (95%). The wetlands had the lowest accuracy about 33%. Both classes of deciduous broadleaf forests and savannas had relatively poor accuracies around 40%. Most errors occurred when shrublands were identified as wetlands, and croplands as deciduous forest. Other errors included confusing deciduous forest with wetlands, and evergreen and deciduous forest with croplands.

B. Data Processing

A subset of the GLCC was prepared for each of 48 continental states. The land-cover distribution in GLCC datasets was computed for each state. Since the classification scheme of GLCC was designed to characterize global land-cover types, class revisions including class removal and merging were necessary for further studying the GLCC datasets for the U.S.

TABLE II
PROPORTION OF LAND-COVER DISTRIBUTION OF NLCD FOR THE STUDIED STATES

Classes \ State	South Dakota	Nevada	West Virginia	Washington	Maine	Florida	Iowa	Wisconsin	Kentucky	Oklahoma
Water	0.02	0.01	0.01	0.02	0.05	0.04	0.01	0.03	0.02	0.02
Low Intensity Residential	<0.01	<0.01	0.01	0.01	0.01	0.05	0.01	0.01	0.01	0.01
High Intensity Residential	<0.01	<0.01	<0.01	<0.01	<0.01	0.02	<0.01	0.01	<0.01	<0.01
Commercial/Industrial/ Transportation	<0.01	<0.01	<0.01	0.01	0.01	0.02	0.02	0.01	0.01	0.01
Bare Rock	0.01	0.05	<0.01	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Quarries/Mines	<0.01	<0.01	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Transitional	<0.01	<0.01	<0.01	0.03	0.02	0.04	<0.01	<0.01	<0.01	<0.01
Deciduous Forest	0.01	<0.01	0.69	0.05	0.23	0.01	0.07	0.31	0.49	0.17
Evergreen Forest	0.03	0.09	0.04	0.4	0.26	0.21	<0.01	0.03	0.03	0.02
Mixed Forest	<0.01	<0.01	0.11	0.04	0.31	0.03	<0.01	0.05	0.08	0.03
Shrubland	0.02	0.79	<0.01	0.17	0.01	0.01	<0.01	<0.01	<0.01	0.04
Orchards/Vineyard	<0.01	<0.01	<0.01	0.01	<0.01	0.03	<0.01	<0.01	<0.01	<0.01
Grasslands/Herbaceous	0.43	0.05	<0.01	0.07	<0.01	0.1	0.05	0.01	<0.01	0.32
Pasture/Hay	0.15	0.01	0.11	0.04	0.01	0.04	0.16	0.21	0.2	0.16
Row Crops	0.24	<0.01	0.02	0.01	0.04	0.09	0.65	0.24	0.13	0.06
Small Grains	0.05	<0.01	<0.01	0.07	<0.01	<0.01	0.01	<0.01	<0.01	0.16
Fallow	0.02	<0.01	<0.01	0.06	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Urban/Recreation Grasses	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	<0.01
Woody Wetlands	<0.01	<0.01	<0.01	<0.01	0.04	0.21	0.01	0.07	0.02	<0.01
Emergent/herbaceous Wetlands	0.02	<0.01	<0.01	<0.01	0.01	0.1	0.01	0.02	<0.01	<0.01

The following GLCC classes were omitted from this study due to small area existing in the U.S.: mixed dryland/irrigated cropland pasture (213), mixed shrubland/grassland (330), deciduous needleaf forest (412), herbaceous wetland (620), and a variety of tundra (810, 820, 830, and 850). The next step was combining some GLCC classes to correspond to the individual NLCD class. For example, the GLCC evergreen broadleaf forest (421) and evergreen needleleaf forest (422) were combined as evergreen forest. The same adjustment was applied to the GLCC dryland cropland/pasture (211) and irrigated cropland/pasture (212) combined as cropland/pasture.

For the rest of GLCC classes, most of them directly corresponded to the NLCD classes except the agricultural classes $(211 + 212, 280, 290)$ and savanna (332). The GLCC classes of cropland and pasture $(211 + 212)$, cropland/grassland mosaic (280), and/woodland mosaic (290) contained more than one land-cover type. The NLCD did not have the class of savanna (332) as the GLCC. This study primarily focused on high-density vegetated lands and urban-related areas; therefore, no further work was applied to the GLCC classes of barren or sparsely vegetated (770), water bodies (500), and snow or ice (900).

A total of 11 GLCC classes were used for the data analyses. They were ten vegetated classes and one urban-related class including grassland, shrubland, deciduous forest, evergreen forest, mixed forest, wooded wetland, cropland and pasture, cropland/grassland mosaic, cropland/woodland mosaic, savanna and urban, and built-up land. The former six classes directly corresponding to the NLCD classes were implemented to evaluate the agreement between the two datasets in 1-km units. The remaining four vegetated classes involving more than one land feature were applied to assess the data relationship. The only urban class was utilized to investigate the relationship of urban-related land-cover between the two datasets. Ten out of the 48 states were chosen for this study, and each was dominated by one of the ten vegetated classes. All of the urban-related lands in ten selected states were analyzed due to a small amount of urban area in each state.

The NLCD for ten states were resampled from 30- to 25-m resolution. The statistical data of land-cover distribution for the ten states was recorded for this study (Table II). A total of 1600 (40×40) small pixels in 25-m resolution constitute one large pixel in 1-km resolution. The land-cover distribution of NLCD within each dominant GLCC class in the correspondent state was calculated in percentage. Moreover, the NLCD distribution within the GLCC urban-related areas was computed for each of the ten states. Only the large pixels completely covered by 1600 small pixels were used for this study. Most edge pixels along each state boundary were discarded due to incomplete information. A database was established to store the land-cover data of GLCC and NLCD for each dominant GLCC class and urban-related areas. Each record of the database represented one $1 \text{ km} \times 1 \text{ km}$ pixel and consisted of the location of the pixel, the land-cover type of GLCC, and the geocorrespondent land-cover composition of NLCD in percentage. The average and standard deviation for each geocorrespondent NLCD class were calculated for each dominant GLCC class and the urban-related areas.

III. RESULTS

A. Agreement Evaluation

The NLCD distributions within the six GLCC classes were analyzed to determine agreements. Each of the six states chosen for the study was dominated by one of the selected land-cover types except Florida. The six states were South Dakota for grassland, Nevada for shrubland, West Virginia for deciduous forest,

TABLE III
PROPORTION OF LAND-COVER DISTRIBUTION OF GLCC FOR THE STUDIED STATES

State / Classes of GLCC	South Dakota	Nevada	West Virginia	Washington	Maine	Florida	Iowa	Wisconsin	Kentucky	Oklahoma
Urban and Built-up Land	<0.01	<0.01	0.01	0.01	<0.01	0.02	0.01	0.01	0.01	0.01
Cropland and Pasture	0.24	<0.01	0.01	0.11	0.01	0.16	**0.92**	0.11	0.02	0.11
Cropland/grassland Mosaic	0.24	<0.01	<0.01	0.02	0.01	0.12	0.05	**0.56**	0.01	0.06
Cropland/woodland Mosaic	<0.01	0	0.03	<0.01	0.01	0.22	<0.01	<0.01	**0.43**	0.04
Grassland	**0.45**	0.05	0	0.14	<0.01	0.04	0.01	<0.01	<0.01	0.33
Shrubland	0.03	**0.79**	0	0.13	<0.01	<0.01	0	<0.01	0	0.01
Savanna	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	**0.29**
Deciduous Forest	<0.01	0.01	**0.9**	0.01	0.26	<0.01	<0.01	<0.01	0.47	0.08
Evergreen Forest	0.03	0.01	<0.01	**0.51**	0.11	0.26	<0.01	<0.01	<0.01	0.03
Mixed Forest	<0.01	<0.01	0.05	0.01	**0.45**	<0.01	0.01	0.16	0.04	0.03
Water Bodies	0.01	<0.01	<0.01	0.06	0.15	0.14	<0.01	0.16	0.01	0.01
Wooded Wetland	0	<0.01	0	0	0	**0.04**	0	0	0	0
Barren or Sparsely Vegetated	0	0.14	0	<0.01	0	<0.01	0	0	0	0
Snow or Ice	0	<0.01	0	<0.01	0	0	0	0	0	0
Total	1	1	1	1	1	1	1	1	1	1

Washington for evergreen forest, Maine for mixed forest, and Florida for wooded wetland (Table III). Although the wooded wetland occupies only 4% of the land-cover in Florida, the state of Florida has the largest wetland in U.S. The averaged results (Table IV) revealed that each studied GLCC pixel at 1-km resolution contained two or more NLCD classes. For example, each studied GLCC grassland pixel on average was occupied by ten NLCD classes including water (1%), bare rock (1%), deciduous forest (1%), evergreen forest (1%), shrubland (4%), grassland (67%), pasture/hay (5%), row crops (8%), small grains (7%), and fallows (4%). The 67% of NLCD pixels within a GLCC grassland pixel agreed as grassland in South Dakota (Table IV). Our results showed that the GLCC datasets in selected states had agreements to the NLCD distribution greater than 60% on average for the classes of grassland, shrubland, deciduous forest, and evergreen forest. Both mixed forest and wooded wetland of GLCC had agreements as low as 33% and 14%, respectively. The mean and standard deviation of each NLCD class simply served as references of NLCD distribution within each of the six studied GLCC classes.

The GLCC grassland in South Dakota dominated more than 45% of land-cover (Table III) and was mostly located in the west [Fig. 1(a)]. Our analyses showed that the agreement was 67% in average for grassland, and the area in disagreement was primarily occupied by shrubs and crops [Fig. 1(b)]. A small proportion of nonvegetated land-cover (water and bare rock) was embedded in the GLCC grassland (Table IV). Both GLCC and NLCD showed that shrubland occupied about 79% of land-cover in Nevada (Tables II and III). Our results exhibited that the GLCC shrubland distribution had a high agreement of 81% to the NLCD shrubland, while the other 19% was mainly distributed by the NLCD evergreen forest and grassland (Table IV). For the forestlands, the agreements on average between the two datasets were about 72% for the deciduous forest in the state of West Virginia and 63% for the evergreen forest in the state of Washington (Table IV). The deciduous forest was mostly confused with mixed forest and pasture in West Virginia, while the evergreen forest was primarily confused with the other types

TABLE IV
SPATIAL COMPARISONS OF LAND-COVER DISTRIBUTION BETWEEN THE GLCC (1 km × 1 km) AND NLCD (25 m × 25 m). THE NUMBERS AND NUMBERS IN PARENTHESES REPRESENTED THE AVERAGED NLCD COMPOSITION AND STANDARD DEVIATION, RESPECTIVELY, FOR THE STUDIED GLCC CLASSES

Classes of GLCC (code) – state / Classes of NLCD (code)	**Grassland** (311) - South Dakota	**Shrubland** (321) - Nevada	**Deciduous Forest** (411) - West Virginia	**Evergreen Forest** (421&422) - Washington	**Mixed Forest** (430) - Maine	**Wooded Wetland** (610) - Florida
Water (11)	0.01 (0.07)	0 (0.03)	0.01 (0.03)	0.02 (0.1)	0.03 (0.11)	0.12 (0.24)
Perennial Ice Snow (12)	0 (0)	0 (0)	0 (0)	0 (0.02)	0 (0)	0 (0)
Low Intensity Residential (21)	0 (0.01)	0 (0.02)	0 (0.03)	0.01 (0.05)	0.01 (0.03)	0.04 (0.12)
High Intensity Residential (22)	0 (0.01)	0 (0.01)	0 (0)	0 (0)	0 (0.01)	0.03 (0.09)
Commercial/Industrial/ Transportation (23)	0 (0.01)	0 (0.02)	0 (0.02)	0 (0.02)	0 (0.02)	0.03 (0.1)
Bare Rock (31)	0.01 (0.05)	0.03 (0.11)	0 (0)	0.02 (0.06)	0 (0)	0 (0.03)
Quarries/Mines (32)	0 (0)	0 (0.02)	0.01 (0.04)	0 (0)	0 (0)	0 (0.04)
Transitional (33)	0 (0)	0 (0)	0.01 (0.03)	0.06 (0.12)	0.03 (0.08)	0 (0.03)
Deciduous Forest (41)	0.01 (0.02)	0 (0.01)	**0.72 (0.21)**	0.08 (0.13)	0.23 (0.2)	0 (0.01)
Evergreen Forest (42)	0.01 (0.06)	0.1 (0.23)	0.03 (0.07)	**0.63 (0.3)**	0.28 (0.21)	0.01 (0.03)
Mixed Forest (43)	0 (0.01)	0 (0.01)	0.1 (0.07)	0.07 (0.1)	**0.33 (0.17)**	0 (0.01)
Shrubland (51)	0.04 (0.1)	**0.81 (0.27)**	0 (0)	0.04 (0.08)	0.01 (0.02)	0 (0.02)
Orchards/Vineyard (61)	0 (0)	0 (0)	0 (0)	0 (0.03)	0 (0)	0 (0.02)
Grasslands/Herbaceous (71)	**0.67 (0.33)**	0.05 (0.11)	0 (0)	0.04 (0.09)	0 (0)	0.1 (0.21)
Pasture/Hay (81)	0.05 (0.15)	0.01 (0.06)	0.1 (0.15)	0.03 (0.11)	0.01 (0.03)	0 (0.03)
Row Crops (82)	0.08 (0.2)	0 (0)	0.02 (0.04)	0 (0.02)	0.03 (0.08)	0.01 (0.08)
Small Grains (83)	0.07 (0.15)	0 (0)	0 (0)	0 (0.02)	0 (0)	0 (0)
Fallow (84)	0.04 (0.11)	0 (0)	0 (0)	0 (0.02)	0 (0)	0 (0)
Urban/Recreational Grasses (85)	0 (0.01)	0 (0)	0 (0)	0 (0.01)	0 (0.02)	0.01 (0.04)
Woody wetlands (91)	0 (0.02)	0 (0)	0 (0.01)	0 (0.02)	0.03 (0.08)	**0.14 (0.2)**
Emergent/Herbaceous Wetlands (92)	0.01 (0.03)	0 (0.02)	0 (0.01)	0 (0.01)	0.01 (0.02)	0.51 (0.37)
Total	1	1	1	1	1	1

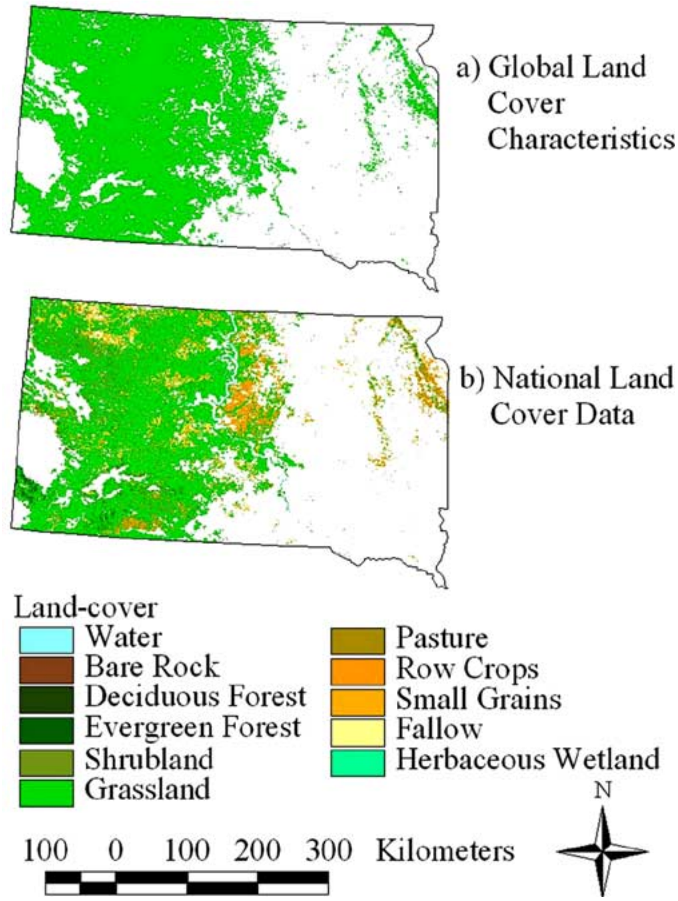

Fig. 1. Grassland distribution in GLCC for (a) the state of South Dakota and (b) the land-cover information in NLCD correspondent to the GLCC grassland in South Dakota.

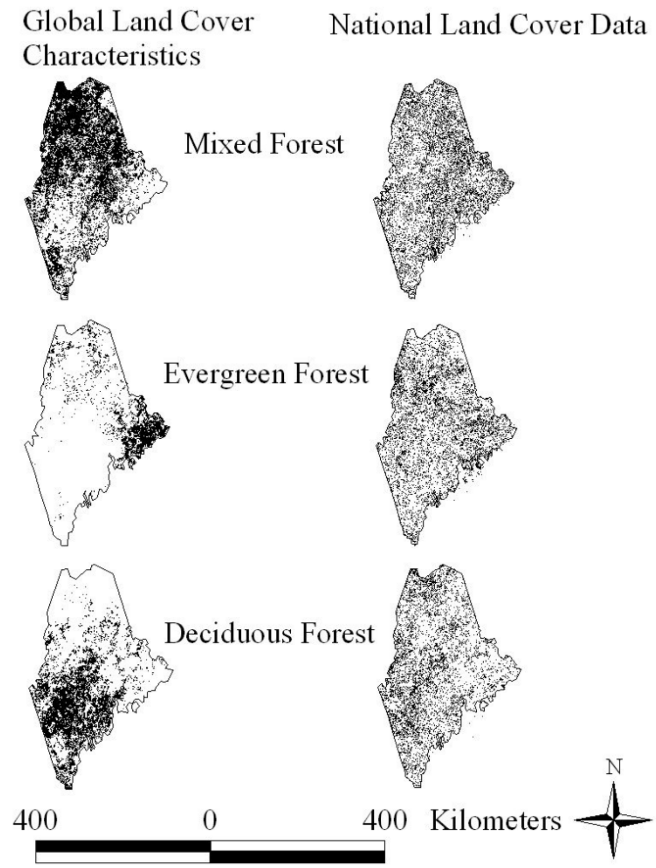

Fig. 2. Forestland distribution for the state of Maine according to the GLCC and NLCD, respectively.

of forests in Washington (Table IV). The classes of wetland and urban were barely related to the above GLCC vegetated land-cover classes.

The previous results showed that the land-cover type of GLCC was resolved based on the major vegetation type in each 1-km^2 pixel. But, the GLCC mixed forest and wooded wetland did not follow the same rule. The GLCC mixed forest covered more than 45% of land in the state of Maine (Table III). Only 33% of NLCD pixels within a GLCC mixed forest pixel agreed as mixed forest in Maine. The results revealed that nearly 51% of GLCC mixed forest in average was actually covered by the NLCD deciduous forest and evergreen forest (Table IV). The GLCC map showed that a high density of mixed forest was located in the north part of Maine, while the NLCD showed that the mixed forest was equally distributed in Maine (Fig. 2). Similar distribution disagreements happened to the evergreen and deciduous forest, which contributed to the poor agreement of mixed forest (Fig. 2). The GLCC wooded wetland in Florida had the lowest agreement around 14%. It was unexpected that more than 50% of GLCC wooded wetland was covered by the NLCD herbaceous wetlands (Table IV), whereas no GLCC herbaceous wetland was found in the U.S. According to the NLCD, the woody wetlands dominated 21% of land-cover in Florida [Table II and Fig. 3(a)], which was much different from the 4% in the GLCC (Table III). Most agreed wooded wetland was located in the southern Florida [Fig. 3(b)]. More than 43% of NLCD woody wetlands were identified as the evergreen forest in GLCC, and another 25% of NLCD woody wetlands were recognized as the cropland/woodland mosaic in GLCC (Fig. 2). Similar situations happened to the NLCD herbaceous wetlands. The NLCD herbaceous wetlands were mostly identified as the cropland/woodland mosaic, cropland and pasture, or wooded wetland in GLCC.

B. Relationship Assessment

Four GLCC classes containing more than one vegetated land-cover type were used for the relation assessment due to each of the four GLCC classes correspondent to two or more NLCD classes. Each of the four states chosen for the study was mainly occupied by one of the selected GLCC classes. The four states were Iowa for cropland and pasture, Wisconsin for cropland/grassland mosaic, Kentucky for cropland/woodland mosaic, and Oklahoma for savanna (Table III). Our results showed that the relationship was strong up to $84\%(16\% + 68\%)$ between the GLCC cropland and pasture and NLCD pasture/hay and row crops in Iowa (Table V). The four GLCC classes were mostly related to the NLCD pasture/hay, row crops, deciduous forest, and grassland (Table V).

The state of Iowa is a key state for corn production and had about 92% of land covered by the GLCC cropland and pasture (Table III). Our analyses showed that the GLCC cropland and pasture were, as expected, primarily related to the NLCD row crops and pasture (Table V). The percentage of NLCD row crops

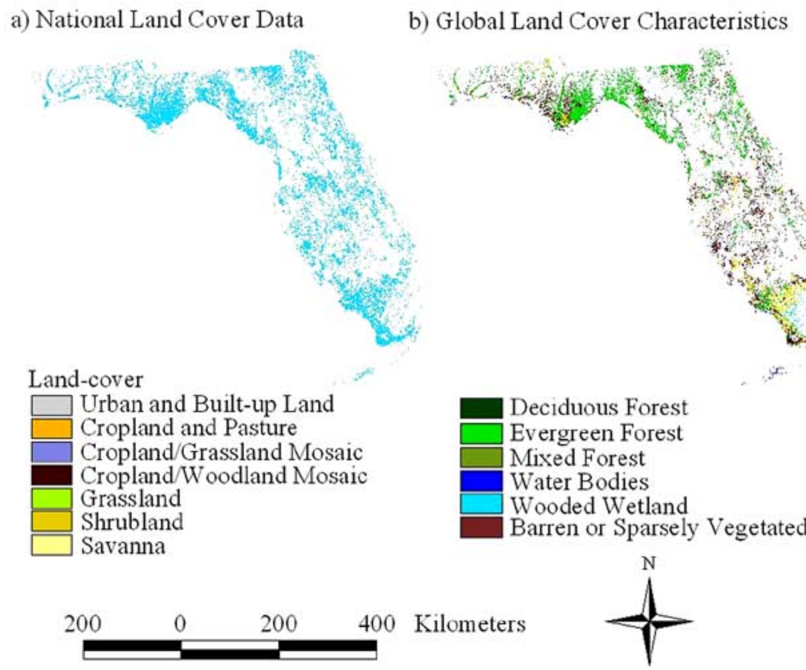

Fig. 3. Woody wetlands distribution in NLCD for (a) the state of Florida and (b) the land-cover information in GLCC correspondent to the NLCD woody wetlands in Florida.

TABLE V
RELATION OF LAND-COVER DISTRIBUTION BETWEEN THE GLCC (1 km × 1 km) AND NLCD (25 m × 25 m). THE NUMBERS AND NUMBERS IN PARENTHESES REPRESENTED THE AVERAGED NLCD COMPOSITION AND STANDARD DEVIATION, RESPECTIVELY, FOR THE STUDIED GLCC CLASSES

Classes of GLCC (code) – state / Classes of NLCD (code)	Cropland & Pasture (211 & 212) - Iowa	Cropland/ Grassland Mosaic (280) - Wisconsin	Cropland/ Woodland Mosaic (290) - Kentucky	Savanna (332) - Oklahoma
Water (11)	0.01 (0.04)	0.02 (0.08)	0.01 (0.05)	0.02 (0.07)
Perennial Ice Snow (12)	0 (0)	0 (0)	0 (0)	0 (0)
Low Intensity Residential (21)	0 (0.04)	0.01 (0.03)	0.01 (0.05)	0.01 (0.06)
High Intensity Residential (22)	0 (0.02)	0 (0.02)	0 (0.01)	0 (0.02)
Commercial/Industrial/ Transportation (23)	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)
Bare Rock (31)	0 (0)	0 (0)	0 (0)	0 (0.01)
Quarries/Mines (32)	0 (0.01)	0 (0.01)	0 (0.02)	0 (0.01)
Transitional (33)	0 (0)	0 (0.02)	0 (0.02)	0 (0.03)
Deciduous Forest (41)	0.06 (0.11)	**0.31 (0.26)**	**0.26 (0.21)**	**0.24 (0.23)**
Evergreen Forest (42)	0 (0)	0.03 (0.06)	0.02 (0.03)	0.02 (0.04)
Mixed Forest (43)	0 (0)	0.04 (0.06)	0.07 (0.06)	0.01 (0.02)
Shrubland (51)	0 (0)	0 (0)	0 (0)	0.02 (0.05)
Orchards/Vineyard (61)	0 (0)	0 (0)	0 (0)	0 (0)
Grasslands/Herbaceous (71)	0.05 (0.06)	0.01 (0.02)	0 (0)	**0.34 (0.26)**
Pasture/Hay (81)	**0.16 (0.17)**	**0.26 (0.22)**	**0.37 (0.22)**	**0.24 (0.24)**
Row Crops (82)	**0.68 (0.28)**	**0.24 (0.21)**	**0.22 (0.19)**	0.04 (0.09)
Small Grains (83)	0.01 (0.02)	0 (0)	0 (0)	0.05 (0.11)
Fallow (84)	0 (0)	0 (0)	0 (0)	0 (0)
Urban/Recreational Grasses (85)	0 (0.02)	0 (0.02)	0.01 (0.03)	0 (0.01)
Woody wetlands (91)	0.01 (0.03)	0.05 (0.12)	0.02 (0.07)	0 (0.03)
Emergent/Herbaceous Wetlands (92)	0.01 (0.03)	0.02 (0.06)	0 (0.01)	0 (0.01)
Total	1	1	1	1

within each GLCC cropland and pasture was greater than 90% in most areas of the northern Iowa except river zones [Fig. 4(a)]. The percentage of NLCD row crops decreased gradually toward the south; meanwhile the density of NLCD pasture/hay increased. The southern central Iowa was mainly occupied by the NLCD pasture/hay [Fig. 4(b)].

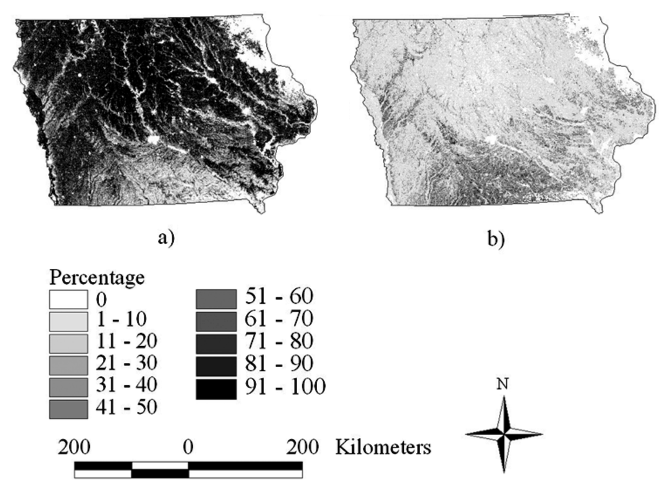

Fig. 4. Percentage of (a) row crops and (b) pasture/hay of NLCD for each 1-km unit of cropland and pasture in GLCC across the state of Iowa.

The GLCC cropland/grassland mosaic and cropland/woodland mosaic were the major land-cover types for the states of Wisconsin and Kentucky, respectively (Table III). Both GLCC land-cover classes had the same relationship to the NLCD deciduous forest, pasture, and row crops (Table V). Both relationships were unexpected. For example, the GLCC cropland/grassland mosaic was strongly related to the NLCD deciduous forest but little to the NLCD grassland (Table V). Another unexpected result was that the GLCC cropland/woodland mosaic had a very high relationship to the NLCD pasture/hay (Table V). Both relationships were possibly caused by the confusions of deciduous forest with croplands and row crops with pasture/hay. The GLCC savanna in central eastern Oklahoma

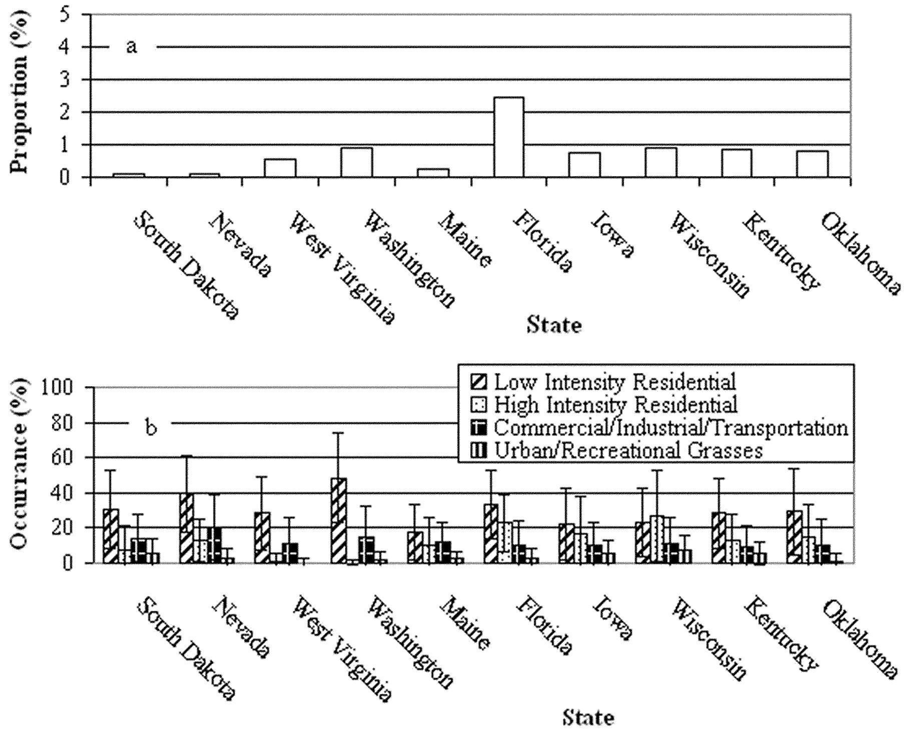

Fig. 5. Proportions in percentage of GLCC urban and built-up land distribution for (a) each of the ten states and (b) the relationship of GLCC urban and built-up land to the NLCD low-intensity residential, high-intensity residential, commercial/industrial/transportation, and urban/recreational grasses on average for each of the ten states. The error bars in (b) represented the stand deviations for each NLCD class.

occupied about 29% of the land-cover (Table III). The results showed that the GLCC savanna was primarily related to the NLCD grassland, and meanwhile properly corresponded to NLCD deciduous forest and pasture (Table V). The NLCD evergreen and mixed forest, shrubland and wetland were barely related to the four GLCC classes. A small proportion of non-vegetated NLCD such as residential, commercial, and water were distributed in the four GLCC vegetated lands (Table V).

C. Urban-Related Area Investigation

Each of the ten states contributed data for studying the relationship of urban-related areas between the two datasets. In general, the GLCC urban and built-up lands occupied a small proportion of land-cover for the ten states ranging from 0.09% (Nevada) to 2.46% (Florida) Fig. 5(a). Our results showed that the GLCC urban and built-up land was strongly correlated to the NLCD low-intensity residential and showed little relationship to the NLCD urban/recreation grasses Fig. 5(b). The relationship between the two datasets was variable state by state for the NLCD high-intensity residential and commercial/industrial/transportation. The GLCC urban and built-up land in the ten states was consistently related to the nonurban NLCD classes such as deciduous forest, grassland, pasture/hay, row crops, and water. The NLCD evergreen forest and mixed forest were occasionally related to the GLCC urban and built-up land as well.

IV. Discussion

A significant source of disagreement for the NLCD was between upland forest classes (deciduous, evergreen, and mixed forests) [8]. The mixed forest was often confused with deciduous forest and evergreen forest in this study. Although both datasets showed that the upland forests occupied around 80% of Maine (Tables II and III), the scenarios of forest distributions were different between the two datasets (Fig. 2). The three forest classes were equally distributed around the state in the NLCD, but visibly centered at three different locations in the GLCC (Fig. 2). The agreement increased from 33% to 84% when three upland forests were grouped as one forest class in Maine (Table IV). Improvements were slight for the classes of deciduous forest from 72% to 85% in West Virginia and evergreen forest from 63% to 78% in Washington (Table IV). The major forests in similar spatial distributions for both GLCC and NLCD explained the smaller degree of improvement for the deciduous and evergreen forests. Meanwhile, data users should be aware that both deciduous and evergreen forest were less confused with mixed forest (Table IV).

A noticeable confusion was found between the wetland classes in GLCC through this study. The NLCD presented that a few of herbaceous wetlands were scattered in several states, but no GLCC pixels in herbaceous wetlands were found in the U.S. The wetland classes in NLCD were classified with accuracy as high as 67% for the wooded wetland and 81% for the herbaceous wetland [8]. This study results showed the herbaceous wetland could not be distinguished from the wooded wetland in GLCC, which could be the result of spectral similarities between the two wetlands. Other land-cover classification schemes for the GLCC tended to combine both wetland classes as one class of permanent wetland, which would promote the agreement result from 14% to 65%.

The similarity level of land-cover percentage and spatial distribution determined the degrees and areas of agreement or relationship between the NLCD and GLCC datasets. Significant spatial differences led to the low agreements for the mixed forest in Maine and wooded wetland in Florida. The high agreement for the shrubland in Nevada was a result that both datasets had very similar numerical (percentage) and spatial (distribution) information (Tables II and III). The NLCD exhibited that $1/3$ of Wisconsin was dominated by deciduous forest, whereas the GLCC showed no deciduous forest in Wisconsin (Tables II and III). The strong disagreement of deciduous forest distribution was one of the factors causing the relationship between the GLCC cropland/grassland mosaic and the NLCD deciduous forest. Although the two datasets had similar percentages (49% versus 47%) for deciduous forest in Kentucky, much of NLCD deciduous forest occurred in areas of GLCC cropland/woodland mosaic. Both datasets showed that agricultural land occupied most of Iowa and had similar spatial distributions. Hence, the GLCC cropland and pasture was correctly correlated to the NLCD row crops and pasture/hay. Land-cover data from other sources are certainly required to validate the areas in low agreements or relationships.

The confusion between the forest and agricultural classes in both NLCD and GLCC datasets was another source causing unanticipated relationships. Three crop-related GLCC classes (cropland and pasture, cropland/grassland mosaic, and cropland/woodland mosaic) were correlated to both pasture/hay and row crops in NLCD. Many croplands in the U.S. were located between pasture/hay and could not be separated out in 1-km pixels. Moreover, the GLCC cropland/grassland mosaic was unexpectedly correlated to the NLCD deciduous forest. This study found that grassland was often confused with agricultural and forest classes. For example, the agricultural classes of NLCD occupied more than 20% of GLCC grassland. The GLCC showed that cropland and pasture covered the entire state of Iowa except the northeast corner dominated by the cropland/grassland mosaic. The NLCD showed the northeast corner of Iowa was occupied by pasture and deciduous forest. The same confusion occurred in Wisconsin where the GLCC cropland/grassland mosaic corresponded to the NLCD pasture/hay and deciduous forest. Similar confusion in Kentucky showed the GLCC cropland/woodland mosaic corresponding to the NLCD pasture/hay.

The urban and built-up lands covered a small area compared to the vegetated lands, but the urban-related information is as critical as vegetation distributions to environmental studies. Several NLCD classes such as deciduous forests, grassland, pasture, and cropland were consistently related to the GLCC urban and built-up lands as a result of the common mixture of urban and vegetated lands. The NLCD class of urban/recreational grasses represented the greenness in urban-related areas; therefore, the relationship was relatively low to the GLCC urban and built-up lands. The relationship between two datasets for the urban-related areas increased up to 55% on average when the NLCD residential and commercial areas were grouped as one urban class for the ten states.

Numerous minor land-cover features were missed in the GLCC datasets due to the 1-km resolution. For example, small grains and fallows are omitted in croplands, and orchards/vineyard could be omitted within either shrublands or croplands. The NLCD classification scheme provided land-cover information in detail, but some minor classes might not be as valuable as others for large-scale research studies. The Anderson Level I classification scheme including water, urban, barren land, forest, agricultural land, wetland, and rangeland could simply eliminate confusion between similar land-cover classes, but might miss some important land features such as shrubland and grassland. More studies are required regarding systematically grouping land-cover classes based on research objectives and data availability. Due to coarse resolution, the total areas of some dominant land features based on the GLCC might be overestimated, while some insignificant land-cover classes were underestimated.

This study assessed the spatial distribution of two land-cover datasets at state scales, and the resulting set of subclass land-cover information for 11 original or derived GLCC classes provided data users the insights of land features in 1-km unit. The subclass information could be changed if the study were confined to a local scale or extended to ecoregional scales. The data agreements and relationships would be significantly improved if the study site had a nearly homogeneous land-cover distribution. For example, the northern Iowa homogeneous croplands could have the data relationship up to 90%. Further work is necessary to compare multiple outcomes of land-cover agreements and relationships between the two datasets in a diversity of scales. Meanwhile, the data applications are critical to approve the importance of this work and to examine the impact of land-cover information at various scales on modeling outputs.

V. Conclusion

This work studied the spatial distributions of the NLCD and GLCC for the classes of agricultural land, forest land, rangeland, wetland, and urban-related areas. The NLCD distribution within each studied GLCC class in 1-km unit was analyzed to determine data similarity in ten U.S. states. Our results showed a general agreement for the classes of grassland, shrubland, deciduous forest, and evergreen forest. Strong disagreement occurred in the classes of mixed forest and wooded wetland. Both NLCD classes of pasture/hay and row crops were indistinguishable in the cropland of GLCC at 1-km resolution. The GLCC classes of cropland and pasture, cropland/woodland mosaic, savanna, and urban and built-up land were appropriately related to the NLCD classes. The confusion between cropland and forest land possibly caused the unusual relationship between the GLCC cropland/grassland mosaic and NLCD deciduous forest. The mixture of urban and vegetated lands caused the constant relationship between the GLCC urban and built-up lands and NLCD vegetated lands. Several minor land features were most likely omitted in the 1-km pixel, and the dominated land-cover class in GLCC was overestimated. Some minor classes in the NLCD were probably redundant for state- to national-scale studies. Hence, a purpose-driven class grouping is necessary for NLCD classes to minimize data confusions and improve data similarity.

One of the strong advantages of satellite data is providing real-time information for research studies. Several updated land-cover maps based on coarse-resolution satellite images have been produced over the last few years, and more will be available for research purposes in the near future. The coarse-resolution data provided less detailed land-cover information and have been restricted for global or national studies so far. The first set of national land-cover datasets in detail was produced based on 1992 data, and the new one using 2000 satellite imagery is still under processing. Near real-time studies certainly need to rely on the updated but less detailed land-cover datasets. This study produced a set of subclass land-cover information in 1-km unit for several GLCC classes, which provided data users the area information of data agreement and relationship between the GLCC and NLCD datasets. Other information such as local land-cover maps derived from fine-resolution satellite data or field survey is required to verify the areas in low agreements or relationships. For those areas in reasonable agreements or relationships, the recently updated global land-cover data could serve as an input for local to regional studies to enhance near real-time environmental research.

REFERENCES

[1] J. E. Vogelmann, S. M. Howard, L. Yang, C. R. Larson, B. K. Wylie, and N. Van Driel, "Completion of the 1990s national land cover dataset for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources," *Photogramm. Eng. Remote Sens.*, vol. 67, pp. 650–662, Jun. 2001.

[2] J. G. Arnold, R. Srinivasan, R. S. Muttiah, and J. R. Williams, "Large area hydrologic modeling and assessment: Part I, model development," *J. Amer. Water Resources Assoc.*, vol. 34, pp. 73–89, Feb. 1998.

[3] R. A. Smith, G. E. Schwarz, and R. B. Alexander, "Regional interpretation of water-quality monitoring data," *Water Resources Res.*, vol. 33, pp. 2781–2798, Dec. 1997.

[4] C. G. Homer, C. Huang, L. Yang, and B. Wylie, "Development of a circa 2000 landcover database for the United States," presented at the *ASPRS Annu. Convention*, 2002.

[5] C. Giri, Z. L. Zhu, and B. Reed, "A comparative analysis of the global land cover 2000 and MODIS land cover datasets," *Remote Sens. Environ.*, vol. 94, pp. 123–132, Jan. 2005.

[6] J. Cihlar, "Thematic Mapper (TM) based accuracy assessment of a land cover product for Canada derived from SPOT VEGETATION (VGT) data," *Can. J. Remote Sens.*, vol. 29, pp. 154–170, Apr. 2003.

[7] J. F. Brown, T. R. Loveland, J. W. Merchant, B. C. Reed, and D. O. Ohlen, "Using multisource data in global land cover characteristics: Concepts, requirements and methods," *Photogramm. Eng. Remote Sens.*, vol. 59, pp. 977–987, Jun. 1993.

[8] L. Yang, S. V. Stehman, J. H. Smith, and J. D. Wickham, "Short communication: Thematic accuracy of MRLC land-cover for the eastern United States," *Remote Sens. Environ.*, vol. 76, pp. 418–422, Jun. 2001.

[9] J. R. Anderson, E. E. Hardy, J. T. Roach, and R. E. Witmer, "A land use and land cover classification system for use with remote sensor data," U.S. Geol. Surv., Reston, VA, USGS Survey Professional Paper 964, 1976.

[10] AVHRR 1-km Global Land 10-Day Composites Project/Campaign Document, USGS. (2005, Apr.). [Online]. Available: http://eosims.cr.usgs.gov:5725/CAMPAIGN_DOCS/avhrr_gc_proj_camp.html.

[11] P. M. Kelly and J. M. White, "Preprocessing remotely sensed data for efficient analysis and classification," in *Proc. SPIE Conf. Applications of Artificial Intelligence 1993: Knowledge-Based Systems in Aerospace and Industry*, 1993, pp. 24–30.

[12] J. C. Eidenshink and J. L. Faundeen, "The 1 km AVHRR global land dataset-first stages in implementation," *Int. J. Remote Sens.*, vol. 15, pp. 3443–3462, Nov. 1994.

[13] S. W. Running, T. R. Loveland, L. L. Pierce, R. R. Nemani, and E. R. Hunt, "A remote sensing based vegetation classification logic for global land cover analysis," *Remote Sens. Environ.*, vol. 51, pp. 39–48, Jan. 1995.

[14] R. Latifovic and I. Olthof, "Accuracy assessment using sub-pixel fractional error matrices of global land-cover products derived from satellite data," *Remote Sens. Environ.*, vol. 90, pp. 153–165, Mar. 2004.

[15] J. Scepan, "Thematic validation of high-resolution global land-cover datasets," *Photogramm. Eng. Remote Sens.*, vol. 65, pp. 1051–1060, Sep. 1999.

Pei-Yu Chen is currently an Assistant Research Scientist at the Blackland Research and Extension Center, part of Texas A&M University system, Temple. Her research interests include image preprocessing, analysis, and application for environmental management and assessment.

Mauro Di Luzio is an Assistant Research Scientist at Blackland Research Center, Texas Agriculture Experimental Station, Texas A&M University, Temple. His research is oriented on ecohydrology. He is currently interested in developing and applying hydrological models for the assessment of the water quality at the watershed and regional scale. His research interest includes the study and development of geographic information systems tools supporting hydrologic, water resource models and concepts, spatial dataset development, and data management for the assessment of agriculture impact on water quality. He has developed GIS/modeling software and tools used in the U.S. (USEPA BASINS tools) and worldwide (AVSWAT) for water quality assessment at the watershed scale.

Jeffrey G. Arnold received the Ph.D. degree in agricultural engineering from Purdue University, West Lafayette, IN, in 1992.

He is currently a Research Leader at the Grassland, Soil, and Water Research Laboratory, U.S. Department of Agriculture's Agricultural Research Service, in Temple, Texas. His research interests include development of hydrologic and water quality models, scaling issues, and national conservation and environmental assessments.